

Article

Resource Optimization for 3D Video SoftCast with Joint Texture/Depth Power Allocation

Saqr Khalil Saeed Thabet ¹, Emmanuel Osei-Mensah ¹, Omar Ahmed ², Abegaz Mohammed Seid ³ and Olusola Bamisile ^{4,*}

¹ School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; saqr-thabet@std.uestc.edu.cn (S.K.S.T.); ecoseimensah@std.uestc.edu.cn (E.O.-M.)

² School of Computer Science and Technology, Hefei University of Technology, Hefei 230009, China; omar@mail.hfut.edu.cn

³ College of Science and Engineering, Hamad Bin Khalifa University, Doha 34110, Qatar; mabegaz@hbku.edu.qa

⁴ College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu 610059, China

* Correspondence: mobb990@gmail.com

Abstract: During wireless video transmission, channel conditions can vary drastically. When the channel fails to support the transmission bit rate, the video quality degrades sharply. A pseudo-analog transmission system such as SoftCast relies on linear operations to achieve a linear quality transition over a wide range of channel conditions. When transmitting 3D videos over SoftCast, the following issues arise: (1) assigning the transmission power to texture and depth maps to obtain the optimal overall quality and (2) handling 3D video data traffic by dropping and re-allocating resources. This paper solves the pseudo-analog transmission resource allocation problem and improves the results by applying the optimal joint power allocation. First, the minimum and the target distortion optimization problems are formulated in terms of a power–bandwidth pair versus distortion. Then, a minimum distortion optimization algorithm iteratively computes all the possible resource allocations to find the optimal allocation based on the minimum distortion. Next, the three-dimensional target distortion problem is divided into two subproblems. In the power-distortion problem, to obtain a target distortion, the algorithm exhaustively solves the closed form of the power resource under a predefined upper-bound bandwidth. For the bandwidth-distortion problem, reaching a target distortion requires solving iteratively for the bandwidth resource closed form, given a predefined power. The proposed resource control scheme shows an improvement in transmission efficiency and resource utilization. At low power usage, the proposed method could achieve a PSNR gain of up to 1.5 dB over SoftCast and even a 1.789 dB gain over a distortion-resource algorithm, using less than 1.4% of the bandwidth.



Citation: Thabet, S.K.S.; Osei-Mensah, E.; Ahmed, O.; Seid, A.M.; Bamisile, O. Resource Optimization for 3D Video SoftCast with Joint Texture/Depth Power Allocation. *Appl. Sci.* **2022**, *12*, 5047. <https://doi.org/10.3390/app12105047>

Academic Editor: David Megías

Received: 29 March 2022

Accepted: 13 May 2022

Published: 17 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: multi-view video plus depth; power allocation ratio; resource allocation; SoftCast

1. Introduction

Owing to the immense growth of industrial and entertainment applications and the rising need for an immersive visual experience, three-dimensional video (3DV) has become a favored choice for many consumers. The 3D scene provides an arbitrary view of the actual scene. Adopting multi-view video plus depth (MVD) [1] to represent 3D video content efficiently has become more common. At the same time, 3D-video-based applications on mobile terminals have gained consumer attention due to the improvement in wireless communication (e.g., 5G and LTE [2]) and the existence of 3D-enabled laptops and naked-eye 3D mobile devices. However, some technical issues remain. Firstly, large amounts of computing resources are required to process a 3D video signal, as the amount of video data

grows proportionally to the number of cameras and the frame depth, which imposes grave challenges for resource-constrained mobile terminals. Secondly, the unpredicted channel condition of the wireless network poses a higher requirement for 3D video transmission than for traditional 2D video. Hence, any proposed efficient wireless transmission method must address these specific issues.

In a 3D video transmission system, the transmitter implements several cameras to record texture videos and depth map frames of the 3D scene. After the receiver selects the preferred virtual viewpoint, the transmitter captures the frames of the adjacent cameras near the selected virtual points. These frames are considered reference viewpoint texture videos and depth maps for that virtual viewpoint. The captured data are source and channel encoded before transmission to the receiver. The demanded virtual viewpoint at the receiver is synthesized from the decoded texture video and depth map frames [3–5] via a technique named depth-image-based rendering (DIBR). Therefore, the virtual view quality is determined by the received texture video quality and depth map quality.

Conventional video transmission over wireless networks relies on Shannon's theorem [6,7], in which source coding is separated from channel coding. Based on this strategy, optimal performance can be obtained provided that the channel condition and capacity are known before source and channel coding, in order to select the proper transmission rate adaptively [8]. However, this strategy does not suit practical wireless communication where the channel condition could vary drastically, and even the receiver is ignorant of the exact channel-state information. The video quality degrades abruptly due to channel capacity mismatch when the channel SNR falls below the acceptable threshold. Even when the channel SNR is high, the reconstruction quality will not improve, as the quantization factor in the digital video compression already determines the upper bound. This phenomenon is known as the *cliff effect* [9].

SoftCast, an uncoded video transmission technique, has been proposed to overcome conventional digital video transmission problems [10–12]. The transmitted signal is transformed into a series of real-value coefficients linearly with the pixel value in the video signal. By skipping quantization and entropy coding, such a transmission becomes lossy in nature. More importantly, the channel noise is directly mapped to the video signal reconstruction error, which makes the reconstruction quality commensurate with the channel SNR of the receiver. Thus, SoftCast tackles the cliff effect. However, when considering SoftCast for transmitting 3DV, several challenges appear.

Firstly, transmitting a large-sized video (i.e., MVD) over a wireless medium consumes substantial amounts of limited resources (power and bandwidth), mostly wasted in carrying reference viewpoints with quite similar content. Secondly, unlike 2D videos, every scene in 3DV consists of texture and depth map frames. Both must be transmitted to the receiver to synthesize the virtual view, which is affected nonlinearly by the received texture videos and the depth maps. Intuitively, assigning more power to texture videos than depth maps will cause an inevitable geometric distortion in virtual views, whereas less power allocation to texture videos would incur additional distortions in reference views. Hence, under limited resources and joint texture/depth power allocation, the problem becomes one of achieving the optimal reconstruction quality. In another case, given a predefined quality at the receiver end, the problem is one of achieving the minimal resource (bandwidth and power) consumption.

This paper proposes incorporating a joint texture–depth power allocation and resource optimization for 3D video transmission, as shown in Figure 1. The optimal power allocation ratio (PAR) between texture and depth is estimated, assuming equal power among all reference viewpoints. Then, the intricate resource control problem of 3D video transmission is considered from the point of view of two optimization problems: minimum distortion and target distortion. Resource allocation algorithms are designed to find the optimal solutions to the formulated problems. The main contributions of this work are as follows:

- The optimal solutions for minimum and target distortion problems are obtained in terms of texture and depth using proposed resource allocation algorithms, and the trade-off relationship between power and bandwidth usage is analyzed.
- The proposed resource control scheme for 3D transmission integrates the optimal PAR with resource allocation algorithms, ensuring balanced power allocation between texture and depth to improve the transmission results.

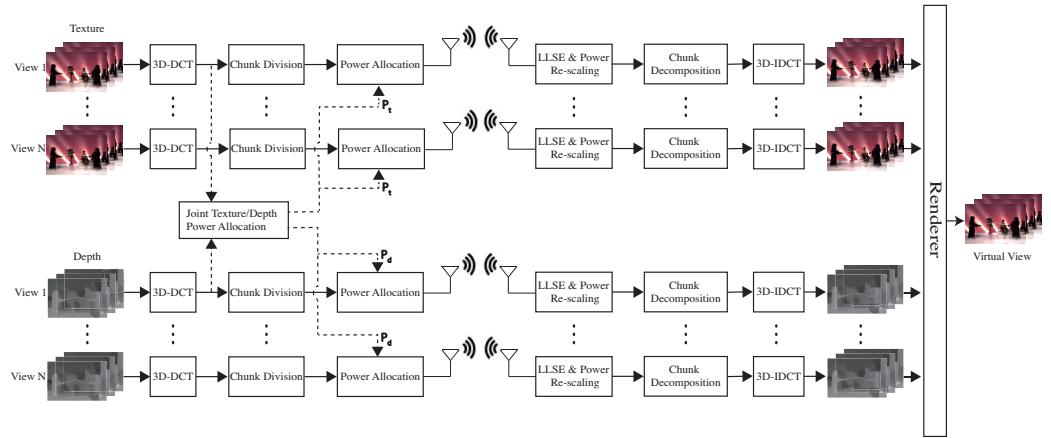


Figure 1. Diagram of the proposed 3D video SoftCast.

The simulation results demonstrate a significant reduction in resource usage in terms of texture and depth, while achieving a satisfactory or desired video quality.

The remainder of the paper is organized as follows. SoftCast is reviewed in Section 2. The minimum distortion optimization is presented in Section 3. Section 4 discusses the target distortion optimization. Sections 5 and 6 show the simulation results and conclude the paper.

2. Related Work

2.1. SoftCast Video Transmission

The energy efficiency of any network shows the extent to which the network can reduce distortion cost efficiently. In [13,14], two examples of improving the network energy efficiency in two different networks are given. Video transmission over wireless network channels is challenging due to the time-varying channel characteristics and power constraints. In such systems, the objective of bit allocation between source and channel is to minimize the power consumption while maintaining satisfactory video quality, and this is also defined as the energy efficiency of the network. The SoftCast transmission scheme is an end-to-end architecture for wireless video transmission that replaces the complicated bit allocation in digital video transmission with a power allocation method. It differs from digital video transmission in the video encoding mechanism, assuring error resilience and transmission. Unlike conventional digital video transmission, the SoftCast approach depends on analog code to achieve linearity and a compression–protection trade-off with an optimal power allocation. In addition to the excellent video quality reception, which is linearly dependent on the channel condition, SoftCast offers scalability that is realized by broadcasting only a single signal. To improve the SoftCast efficiency, Fan et al. [15] first replaced 3D-DCT with motion-compensated temporal filtering (MCTF) and then proposed DCast [16], a distributed source coding that exploits the inter-frame redundancy. In [17,18], adaptive chunk division was incorporated to achieve optimal transmission power usage, and in [19], the bandwidth and computation requirement were reduced. Some other studies suggested a hybrid digital-analog framework, incorporating analog transmission alongside digital transmission to take advantage of both systems, achieving a balance between robust adaptation capability and high coding efficiency [15,20–24]. These systems can be categorized based on the type of information to be digitized and how the analog and

digital parts share the channel. With the increase in the wireless traffic burden on wireless networks, efficient network resource exploitation becomes necessary. Based on the Shannon capacity [6], the network capacity is enhanced via bandwidth increase and frequency reuse. In HDA-SIM [25], the bandwidth of digital traffic is treated as a hidden resource, where the analog modulated data are superimposed over existing digital traffic. He et al. [26] proposed MCast, a linear video transmission system that retransmits data across multiple time slots and multiple channels to exploit the time and frequency resources. These authors also proposed MUcast [27] to solve the problem of resource allocation to realize efficient multi-user video transmission.

To improve the network's energy efficiency in pseudo-analog 3D video transmission. Yang et al. [28] suggested an uncoded wireless depth map transmission scheme using mean-removed block-based DCT to improve decorrelation and view-synthesis distortion, to enhance energy allocation efficiency. Power allocation scaling is performed on rearranged chunks that include inter-block DCT coefficients. Luo et al. [29] considered the view-synthesis distortion in the power-distortion optimization problem to achieve optimal reconstruction quality in the reference and virtual views. The optimization problem was then solved in a closed form of texture/depth power allocation.

In general, all the studies mentioned followed the minimum distortion optimization approach for a pseudo-analog system. Only a few studies addressed the target distortion. In [30], Liu et al. proposed prediction models to solve the target distortion optimization problem for single-view video. The curve-fitting-based resource control algorithm allocates the constrained bandwidth and power resources to obtain the predefined quality. Zhang et al. [31] developed resource allocation algorithms for 3D video. In distortion-resource optimization, the DR algorithm exhaustively searches all possible discrete power and channel resources. For resource-distortion optimization, the optimization problem is divided into power and channel optimization subproblems. Then, considering fixing one resource each time given a desired distortion, the RD algorithm looks for the optimal solution after searching in an exhaustive manner. Nonetheless, it combines texture and depth using 5D-DCT without considering the characteristic distinction. Unlike 2D video pseudo-analog transmission systems, which resort to discarding chunks with the smallest variation to meet the constrained resource requirement without significantly degrading quality, 3D video pseudo-analog transmission systems can face some complications. Due to the involvement of synthesis distortion, the source data that can be discarded from both texture and depth frames must significantly degrade neither the virtual view quality nor the reference view quality, rather than only the latter. Hence, this work aims to perform resource allocation while jointly balancing the power assigned to texture and depth maps to avoid causing geometric distortion in virtual views or additional distortions in reference views, to improve resource allocation in 3D video transmission.

2.2. SoftCast Power Allocation

In a conventional wireless video transmission scheme, the real-value coefficients are transformed into a bitstream for transmission, and channel coding is adopted to add parity bits to cope with drastic channel noise. This strategy demolishes the numeric properties within the video data and leads to the cliff effect. SoftCast avoids this by transmitting the real-value coefficients directly without bitstream transformation. In SoftCast, a novel error protection scheme was developed by scaling the magnitude of the transmitted coefficient. For a specific conventional 2D video sequence, consider applying 3D-DCT to every GOP. The resulting stream of DCT coefficients is divided into N chunks x_1, x_2, \dots, x_N with size hxw , hence $x_i[j], j = 1, \dots, (hxw)$. The amount of information in each chunk is captured by its entropy (i.e., average energy) λ_i . Each chunk x_i is scaled up to $y_i[j] = g_i * x_i[j]$, $g_i > 1$ at the sender. The values y are then transmitted over the wireless channel. The channel noise is taken to be AWGN with zero mean and variance σ_n^2 . The received value becomes

$\hat{y}_i[j] = y_i[j] + n_i[j]$. To estimate the reconstructed signal $\hat{x}_i[j]$ given $\hat{y}_i[j]$, the linear least squares estimator (LLSE) is used:

$$\hat{x}_i[j] = \frac{\lambda_i g_i}{\lambda_i g_i^2 + \sigma_n^2} \hat{y}_i[j] \quad (1)$$

Consider how SNR changes the LLSE. At high SNR [11], the reconstructed signal $\hat{x}_i[j]$ is estimated by simply scaling down the received signal, $\hat{x}_i[j] = \hat{y}_i[j]/g_i$. For the scale up/down scheme, the reconstruction error is only σ_n^2/g_i^2 compared to σ_n^2 for the direct transmission error. Therefore, this scheme remarkably enhances the transmission error protection when g_i is relatively large. Since the power budget for any transmission system is a constrained resource, scaling up some signal samples with more power means leaving other signal samples with less power. Therefore, power allocation is conducted to determine the optimal scaling factors for minimizing the total reconstruction distortion for the entire transmitted sequence. In [11,18,32], the optimal factor for scaling chunk i is:

$$g_i = \lambda_i^{-1/4} \left(\sqrt{\frac{P}{(hxw) \sum_{i=1}^N \sqrt{\lambda_i}}} \right) \quad (2)$$

where P is the total power budget. The total reconstruction distortion under optimal power allocation is

$$D = (hxw) \sum_{i=1}^N \frac{\sigma_n^2}{g_i^2} = \frac{(hxw)^2 \sigma_n^2}{P} \left(\sum_{i=1}^N \sqrt{\lambda_i} \right)^2 \quad (3)$$

From (2), it becomes clear that for traditional 2D video, the reconstruction distortion D is inversely related to the total power P . The value of D can be computed once P is known. However, for 3D video, the power budget P is shared by texture videos and depth maps, which together determine the overall quality of the received 3D video. Hence, to optimally allocate the total power between texture videos and depth maps, the texture/depth power allocation ratio is estimated, considering the texture/depth distortion trade-off [29].

2.3. Resource Allocation Optimization Problem

Optimization problems for the power–bandwidth pair versus distortion address resource constraint challenges in 3D video pseudo-analog transmission. For example, the objective of minimum distortion optimization is to minimize the overall distortion for a given resource usage:

$$\begin{aligned} \min & \quad D \\ \text{s.t.} & \quad R < R_{available} \end{aligned} \quad (4)$$

where D is the distortion and $R_{available}$ represents the available power and bandwidth resources. The target distortion optimization minimizes the resource usage for a given distortion:

$$\begin{aligned} \min & \quad R \\ \text{s.t.} & \quad D < D_{exp} \end{aligned} \quad (5)$$

where D_{exp} is the expected distortion. It must be noted that as pseudo-analog systems transform video data into independent chunks, which means each chunk contributes independent distortion to the entire video quality, the distortion of a chunk caused by a transmission error does not affect any other chunk. Hence, discarding chunks to fit the bandwidth does not prevent each receiver from obtaining a video quality proportional to its channel conditions, and due to the heavy data traffic involved in 3D video transmission, it is crucial to optimally select chunks for transmission and reuse the saved power efficiently.

In summary, the problem becomes: (a) the optimal allocation of the total power between texture videos and depth maps and (b) the joint optimal resource (power and bandwidth) allocation for the entire signal sequence to obtain satisfactory video quality.

3. Minimum Distortion Optimization

3.1. Investigation and Motivation

In pseudo-analog transmission, every scaled DCT chunk is given its own slot in the transmission channel. An advantage of the compacting nature of DCT is that low variance coefficients dominate the high spatial frequencies. Chunks with the lowest variance are considered the least important chunks, denoted LP (low-priority) because they contribute least to the video quality. Therefore, discarding these chunks to meet the constrained resource requirement will not lead to a significant quality degradation. The power saved from chunk dropping is re-allocated to other more important HP (high-priority) chunks for efficient resource utilization, particularly in stringent power-constraint conditions. Therefore, we have a joint problem of power allocation and chunk selection for 3D video transmission.

3.2. Problem Formulation

Given a 3D video sequence, a 3D-DCT is used to decorrelate the texture and depth frames of the reference viewpoints in one GOP, which results in N chunks of DCT coefficients with variance indicated by $\lambda_1, \lambda_2, \dots, \lambda_N$. Without loss of generality, assume $\lambda_i \geq \lambda_j$ for all $i < j$. Let K be the available bandwidth resources (e.g., frequency or time slots). Let $k = [k_1, k_2, \dots, k_N]$ be a binary allocation vector for each GOP. Therefore, the chunk selection is indicated as follows: chunk i is discarded when $k_i = 0$ and transmitted over a bandwidth resource slot when $k_i = 1$. The expected MSE of the i th chunk is denoted as the Euclidean norm $\varepsilon_i = \|x_i[j] - \hat{x}_i[j]\|^2$. The optimal values of power allocation $\rho^* = [\rho_1^*, \rho_2^*, \dots, \rho_N^*]$ and bandwidth allocation are found by minimizing the transmission distortion under constrained power and bandwidth resources. The problem is formulated as follows [33]:

$$MSE(k, \rho) = \min_{(k, \rho)} \mathbf{E}[\varepsilon_i | \sigma_i^2] = \frac{1}{N} \sum_{i=1}^N \frac{\lambda_i}{k_i \rho_i + 1} \quad (6)$$

$$\text{s.t. } \sum_{i=1}^N k_i < K \quad (7)$$

$$\frac{1}{N} \sum_{i=1}^N k_i \rho_i \leq \frac{P}{N(h \times w) \sigma_n^2} := SNR_{chk} \quad (8)$$

$$k_i \in \{0, 1\}, \quad 1 \leq i \leq N \quad (9)$$

$$\rho_i \geq 0, \quad 1 \leq i \leq N \quad (10)$$

where $\rho_i = \frac{g_i^2 \lambda_i}{\sigma_n^2}$ is the signal noise-to-power ratio for the i th chunk for a given power allocation, K should be in the range $[0, N]$, and SNR_{chk} is the SNR budget for each chunk. As systems run short on bandwidth, with $K < N$, to minimize distortion the system will resort to transmitting a number K of HP chunks and discarding the remaining chunks.

$$k = \begin{cases} 1, & \text{if } i \leq K \\ 0, & \text{if } i > K \end{cases} \quad (11)$$

By re-expressing MSE as

$$MSE = \frac{1}{N} \left(\sum_{i=1}^K \frac{\lambda_i}{\rho_i + 1} + \sum_{i=K+1}^N \lambda_i \right) \quad (12)$$

the problem becomes one of finding the optimal power allocation ρ_i^* and bandwidth K^* . Since the second-order derivative of the objective function in (12) is derived as

$$\frac{\partial^2 \text{MSE}}{\partial \rho_i^2} = \frac{2}{N} \sum_{i=1}^K \frac{\lambda_i}{(\rho_i + 1)^3} > 0, \rho_i > 0 \quad (13)$$

the Lagrange multiplier method is used to solve (12), after defining $\mu > 0$ and the Lagrange function J as

$$\min_{(k, \rho)} J = \frac{1}{N} \left(\sum_{i=1}^K \frac{\lambda_i}{\rho_i + 1} + \sum_{i=K+1}^N \lambda_i \right) - \mu \left(\frac{1}{N} \sum_{i=1}^K \rho_i - SNR_{chk} \right) \quad (14)$$

By setting $\frac{\partial J}{\partial \rho_i} = 0, i = 1, 2, \dots, N$ and $\frac{\partial J}{\partial \mu} = 0$

$$\begin{cases} \frac{\partial J}{\partial \rho_i} = \frac{1}{N} \frac{-\lambda_i}{(\rho_i + 1)^2} + \frac{\mu}{N} = 0 \\ \frac{\partial J}{\partial \mu} = \frac{1}{N} \sum_{i=1}^K \rho_i - SNR_{chk} = 0 \end{cases} \quad (15)$$

$$SNR_{chk} = \frac{1}{N} \left(\frac{\sum_{i=1}^K \sqrt{\lambda_i}}{\frac{\sqrt{\lambda_i}}{(\rho_i + 1)}} - K \right) = \frac{\sum_{i=1}^K \sqrt{\lambda_i}}{N \cdot \sqrt{\lambda_i}} (\rho_i + 1) - \frac{K}{N} \quad (16)$$

The optimal power allocation value ρ_i^* is derived as

$$\rho_i^* = \frac{N \sqrt{\lambda_i}}{\sum_{i=1}^K \sqrt{\lambda_i}} \left(SNR_{chk} + \frac{K}{N} \right) - 1 \approx \frac{N \sqrt{\lambda_i}}{\sum_{i=1}^K \sqrt{\lambda_i}} SNR_{chk} \quad (17)$$

This confirms the chunk scaling in (2).

$$g_i^* = \sqrt{\frac{\sigma_n^2 \cdot \rho_i}{\lambda_i}} = \sqrt{\frac{\sigma_n^2 \cdot N \cdot SNR_{chk}}{\sqrt{\lambda_i} \sum_{i=1}^K \sqrt{\lambda_i}}} = \lambda_i^{-1/4} \sqrt{\frac{P}{(hxw) \sum_{i=1}^K \sqrt{\lambda_i}}} \quad (18)$$

From (12), the total power budget P is inversely proportional to the reconstruction distortion. Increasing the power budget increases ρ_i^* and therefore decreases the distortion. Nevertheless, the bandwidth resource K in (17) is the summation term upper bound. Its discrete nature complicates the effort to attain the optimal bandwidth K^* in closed form.

3.3. Problem Solution

We propose using a greedy search algorithm to find the optimal bandwidth K^* , given that the transmitter has complete knowledge of the chunk variance λ_i and the available transmission power P . The transmitter exhaustively searches all the possible discrete bandwidth resources to reach the number of chunks that leads to distortion minimization. With the optimal chunk selection, the amount of traffic is reduced without critical performance degradation, and users' experience improves as the latency is greatly shortened. Meanwhile, other users in the network may utilize the saved bandwidth resources.

In Algorithm 1, after computing the average power of the chunks in each GOP, we calculate all the possible power and bandwidth allocations throughout the GOP in an exhaustive manner. Then, at line 11, the maximum PSNR is chosen to determine the optimal bandwidth usage K^* , the corresponding optimal chunk selection k_i^* , and the optimal power allocation ρ_i^* . In some cases, the performance curve tends to flatten after a certain range, as shown in Figure 2, indicating that an excessive number of channels are used to accomplish a marginal improvement in PSNR, which is apparently ineffective. Therefore, a control parameter τ in the range $[0, 1]$ is introduced in line 14 to find a suboptimal solution. The bandwidth resource usage can be significantly decreased by slightly

sacrificing the PSNR performance. Note that, as 3D-DCT decorrelation transformation is implemented in this scheme, the greedy search in each GOP must be conducted on each reference viewpoint's texture and depth map separately. When conducting minimum distortion optimization under the optimal PAR, the power allocation SNR_{chk} in (17) changes, depending on whether it is assigned to a texture or a depth map. Hence, a different optimal bandwidth usage K^* is chosen, leading to selection of another set of optimal chunks $k_i^{c^*}$ and optimal power allocations $\rho_i^{c^*}$.

Algorithm 1: Minimum Distortion Optimization

Input: $N, \tau, c = [\text{texture}, \text{depth}]$
Output: $k_i^{c^*}, \rho_i^{c^*}, K^{c^*}$
initialization: $\lambda_i = 0, SNR_{chk}^t = SNR_{chk} \cdot \left(\frac{1}{(1+PAR)} \right), SNR_{chk}^d = SNR_{chk} - SNR_{chk}^t;$

- 1: **for** $i = 1$ **to** N
- 2: Compute chunk variance λ_i^t and λ_i^d
- 3: **end**
- 4: **for** $c = 1$ **to** 2
- 5: **for** $n = 1$ **to** N
- 6: $K \leftarrow n;$
- 7: Calculate optimal power allocation $\rho_{i,n}^c \approx \frac{N\sqrt{\lambda_i^c}}{\sum_{i=1}^K \sqrt{\lambda_i^c}} SNR_{chk}^c$
- 8: Calculate MSE_n^c via (12).
- 9: Obtain $PSNR_n^c = 10\log_{10}\left(\frac{(256-1)^2}{MSE_n^c}\right)$
- 10: **end**
- 11: $PSNR_{max}^c = \max_{n \in \{1,2,\dots,N\}} PSNR_n^c.$
- 12: $K_{max}^c \leftarrow n;$
- 13: **for** $n = 1$ **to** N
- 14: **If** $PSNR_n^c > \tau \cdot PSNR_{max}^c$
- 15: $K_{\tau \cdot max}^c \leftarrow n;$
- 16: **break**;
- 17: **end**
- 18: **end**
- 19: Calculate $k_i^{c^*}, \rho_i^{c^*}$ using K^{c^*} as in (11) and (17);
- 20: **end**

4. Target Distortion Optimization

4.1. Investigation and Motivation

Each GOP may have a different compressibility level based on the video sequence content (e.g., rate of HP chunks and LP chunks). Given a limited power and bandwidth usage for a video sequence, some GOPs will be more distorted than others under the minimum distortion optimization. Thus, the distortion fluctuations over GOPs reflect a poor viewing experience (even though the overall PSNR might be maximized). As the channel SNR increases, the quality distortion decreases, given the same limiting constraints for that constrained optimization. Nevertheless, this change in quality is visually indistinguishable by the human eye if the PSNR value is high. These observations show that maintaining a relatively more stable distortion over GOPs is desirable. Thus, for high SNR channels, instead of allowing the PSNR to increase constantly, we keep it at a certain high value. At the same time, other users can use the saved constrained resources. Hence, the problem becomes one of reaching a balanced combination of power allocation and chunk selection for a target distortion.

4.2. Problem Formulation

Since the target distortion optimization comprises bandwidth, power, and distortion, it is considered a three-dimensional problem. As will be demonstrated later, bandwidth and power constraints are exchangeable. Consequently, the optimization problem (5) is

decomposed into two two-dimensional subproblems, where one constraint is fixed in each subproblem. For instance, we formulate the minimization of power resource usage given a distortion constraint \overline{MSE} and a bandwidth resource usage K as a power distortion optimization problem:

$$\begin{aligned} SNR_{chk}(\rho) = \arg \min_{\rho} & \frac{1}{N} \sum_{i=1}^r \rho_i \\ \text{s.t.} & \frac{1}{N} \left(\sum_{i=1}^r \frac{\lambda_i}{\rho_i + 1} + \sum_{i=r+1}^N \lambda_i \right) \leq \overline{MSE} \\ & 1 \leq r \leq K, \quad r \in \mathbb{Z} \end{aligned} \quad (19)$$

On the other hand, we also formulate the minimization of bandwidth resource usage given a target distortion \overline{MSE} and a power budget SNR_{chk} as a bandwidth distortion optimization problem:

$$\begin{aligned} K(\rho) = \arg \min_{\rho} & K \\ \text{s.t.} & \frac{1}{N} \left(\sum_{i=1}^K \frac{\lambda_i}{\rho_i + 1} + \sum_{i=K+1}^N \lambda_i \right) \leq \overline{MSE} \\ & \frac{1}{N} \sum_{i=1}^K \rho_i \leq SNR_{chk} \end{aligned} \quad (20)$$

4.3. Problem Solution

Since the power scaling factor g_i is continuous, while the bandwidth usage variable K is discrete, these two subproblems are considered mixed-integer nonlinear programming (MINLP) problems that are NP-hard to solve optimally. Therefore, greedy search algorithms are proposed to approximate the resource allocation for power and bandwidth distortion problems.

4.3.1. The Power Distortion Optimization

The minimal power use SNR_{chk} in (19) is found by searching exhaustively all the feasible $r \in [1, K]$. For each fixed bandwidth resource usage r , the subproblem is solved as

$$SNR_{chk}(\rho) = \arg \min_{\rho} \frac{1}{N} \sum_{i=1}^r \rho_i \quad (21)$$

$$\text{s.t.} \frac{1}{N} \left(\sum_{i=1}^r \frac{\lambda_i}{\rho_i + 1} + \sum_{i=r+1}^N \lambda_i \right) \leq \overline{MSE} \quad (22)$$

$$\rho_i \geq 0 \quad (23)$$

To use the Lagrange multiplier method to solve this problem, we define a Lagrange multiplier $\mu > 0$, then

$$\min_{\rho} J = \sum_{i=1}^r \rho_i + \mu \left(\sum_{i=1}^r \frac{\lambda_i}{\rho_i + 1} + \sum_{i=r+1}^N \lambda_i - N \cdot \overline{MSE} \right) \quad (24)$$

After setting $\frac{\partial J}{\partial \rho_i} = 0, i = 0, 1, 2, \dots, r$ and $\frac{\partial J}{\partial \mu} = 0$, the optimal solution is calculated by solving

$$\begin{cases} \frac{\partial J}{\partial \rho_i} = 1 + \mu \frac{-\lambda_i}{(\rho_i + 1)^2} = 0 \\ \frac{\partial J}{\partial \mu} = \sum_{i=1}^r \frac{\lambda_i}{(\rho_i + 1)} + \sum_{i=r+1}^N \lambda_i - N \cdot \overline{MSE} = 0 \end{cases} \quad (25)$$

The optimal power allocation value ρ_i^* is derived as

$$\rho_i^* = \begin{cases} \frac{\frac{1}{N} \cdot \sqrt{\lambda_i} \sum_{j=1}^r \sqrt{\lambda_j}}{\overline{MSE} - \frac{1}{N} \sum_{j=r+1}^N \lambda_j} - 1, & i = 1, 2, \dots, r \\ 0, & i = r + 1, \dots, N \end{cases} \quad (26)$$

The algorithm procedures for power distortion optimization demonstrated in Algorithm 2 are performed on a per GOP basis. Throughout the exhaustive search, the optimal power values lie in three regions. At first, $\rho_i \leq 0$, as the number of channels is still insufficient. Subsequently, the distortion caused by discarded chunks gradually lessens, and when $\overline{MSE} - \frac{1}{N} \sum_{i=r+1}^N \lambda_i > 0$, the optimal power will start to comply with the constraint (23). This feasible solution continues for a certain range of r . Finally, beyond this range, up to $r = K$, we have $\rho_i \leq 0$. Therefore, it is better to stop the exhaustive search iteration, as in lines 9 to 11. In line 7, the objective function of problem (19) is calculated in a closed form. Finally, among all the SNR_{chk} s, the one with the minimum value is chosen, and the associated power allocation ρ_i^* is adopted.

Algorithm 2: Power Distortion Optimization

```

Input:  $N, K, \overline{MSE}, c = [\text{texture}, \text{depth}]$ 
Output:  $SNR_{chk}^c, \rho_i^{c*}$ 
initialization:  $\lambda_i = 0$ 
1: for  $i = 1$  to  $N$ 
2:   Compute chunk variance  $\lambda_i^t$  and  $\lambda_i^d$ 
3: end
4: for  $c = 1$  to 2
5:   for  $r = 1$  to  $K^c$ 
6:      $K^c \leftarrow n$ ;
7:     Calculate optimal power allocation  $\rho_{i,r}^c$  via (26)
8:     Calculate  $SNR_{chk,r}^c$  via (21).
9:     If  $(\overline{MSE} - \frac{1}{N} \sum_{i=r+1}^N \lambda_i^c) > 0$  and  $\rho_{i,r}^c < 0$ 
10:       break;
11:     end
12:   end
13:    $SNR_{chk}^{c*} = \min_{r \in \{1, 2, \dots, K-1\}} SNR_{chk}^c$ 
14: end

```

4.3.2. The Bandwidth Distortion Optimization

We exhaustively search the values of K in ascending order for the minimal bandwidth use K . The subproblem is solved for each K as follows:

$$MSE(\rho) = \arg \min_{\rho} \frac{1}{N} \left(\sum_{i=1}^K \frac{\lambda_i}{\rho_i + 1} + \sum_{i=K+1}^N \lambda_i \right) \quad (27)$$

$$\text{s.t. } \frac{1}{N} \sum_{i=1}^K \rho_i \leq SNR_{chk} \quad (28)$$

Using the Lagrange multiplier technique, the optimal power ρ_i^* is derived as

$$\rho_i^* \approx \frac{N \cdot \sqrt{\lambda_i}}{\sum_{j=1}^K \sqrt{\lambda_j}} SNR_{chk} \quad (29)$$

In Algorithm 3, under a given distortion \overline{MSE} and power budget SNR_{chk} , the corresponding K is infeasible if the objective value (27) is greater than the expected distortion. The search continues by increasing r by 1 and solving the problem (27) until the search reaches a feasible K . It is also possible that an exhaustive search for a particular GOP does not reach a feasible K , which means the expected distortion requirement is too high for a specific SNR_{chk} or the power budget SNR_{chk} is insufficient to improve transmission up to a particular quality. In both cases, all N channels are used. Hence, more resources must be allocated. Here, it is suggested that the power budget is increased with a 0.01 step size until a value is reached that will lead to a feasible K .

Algorithm 3: Bandwidth Distortion Optimization

```

Input:  $N, SNR_{chk}^t, SNR_{chk}^d, \overline{MSE}, c = [\text{texture, depth}]$ 
Output:  $K^{c^*}, \rho_i^{c^*}$ 
initialization:  $\lambda_i = 0, r = 0, MSE_r = \infty$ 
1: for  $i = 1$  to  $N$ 
2:   Compute chunk variance  $\lambda_i^t$  and  $\lambda_i^d$ 
3: end
4: for  $c = 1$  to 2
5:   while  $MSE_r^c > \overline{MSE}$  do
6:      $r = r + 1$ 
7:     Calculate optimal power allocation  $\rho_{i,r}^c$  via (29)
8:     Calculate  $MSE_r^c$  via (27)
9:   end
10:   $K^{c^*} \leftarrow r, \rho_i^{c^*} \leftarrow \rho_{i,r}^c$ , for all  $i$ 
11: end
  
```

4.3.3. Power and Bandwidth Trade-Off

To achieve a desirable video quality, either the power consumption is optimized under a constrained bandwidth usage or the bandwidth usage is optimized under a limited power budget. Each point of the resultant trade-off curve between power and bandwidth represents a possible power and bandwidth usage combination that attains the same video quality. In multi-user systems, each viewer has a different power and bandwidth resource budget, where a joint optimization that relies on choosing a proper power and bandwidth usage combination can be implemented to preserve resource consumption.

5. Simulation Results

Several simulation experiments were conducted to assess the performance of the proposed uncoded 3D video wireless transmission.

Test Sequence: For the multi-view plus depth video datasets, different standard reference 3D video sequences were considered: Kendo, Balloons [34], Newspaper provided by GIST, South Korea, and PoznanStreet, PoznanHall2 [35], Dancer, gtFly provided by Nokia, Finland, and Shark by NICT, Japan. The standard video sequences Dancer, gtFly, and Shark are computer-generated 3D scenes with ground-truth depth maps, whereas the content of the other 3D videos consist of captured real 3D scenes. The tested sequence configurations are listed in Table 1.

Parameter settings: The GOP size was set to eight frames. Depending on the sequence resolution in Table 1, each frame was divided into $16 \times 16 = 256$ chunks. As mentioned previously, virtual viewpoint synthesis requires transmitting the texture and depth frames of the two adjacent cameras near the selected virtual point. Hence, one GOP consists of

$2 \times 2 \times 8 \times 256 = 8192$ chunks. The 3D-HEVC Test Model (HTM) v16.3 software [36] was used to synthesize the virtual viewpoints. MATLAB was used to conduct the simulation experiments, where the wireless transmission parameters were based on the 802.11 a/g standard. In addition, the wireless transmission experiments were investigated over AWGN-based channels.

Metric: For video quality assessment, although the proposed algorithms were designed based on the objective performance metric PSNR, the perceptual metric SSIM still revealed important performance evaluation aspects.

Table 1. 3D video sequences used in the experiments [29].

Sequence	3D Scene Nature	Reference Viewpoints	Synthesized Viewpoints	Resolution
Balloons	real	1 and 5	2, 3, 4	1024×768
Kendo	real	1 and 5	2, 3, 4	1024×768
Newspaper	real	2 and 5	2.5, 3, 3.5	1024×768
Dancer	computer-generated	1 and 5	2, 3, 4	1920×1088
gtFly	computer-generated	1 and 5	2, 3, 4	1920×1088
PoznanHall2	real	5 and 7	5.5, 6, 6.5	1920×1088
PoznanStreet	real	3 and 5	3.5, 4, 4.5	1920×1088
Shark	computer-generated	1 and 5	2, 3, 4	1920×1088

5.1. Performance Evaluation for PAR

The optimal PAR was estimated based on the joint texture/depth power allocation method [29]:

$$\text{PAR} = \frac{\sqrt{M + \sum_{i=1}^L \alpha_i} \sqrt{\sum_{m=1}^M (\sum_{j=1}^N \sqrt{\lambda_{m,j}^t})^2}}{\sqrt{\sum_{i=1}^L \beta_i} \sqrt{\sum_{m=1}^M (\sum_{j=1}^N \sqrt{\lambda_{m,j}^d})^2}} \quad (30)$$

where M is the number of reference viewpoints transmitted in a 3D video SoftCast transmission, L is the number of virtual viewpoints to be synthesized at the receiver, t and d represent the original texture video and depth map, respectively, and α_i and β_i represent the parameters of the distortion model of both the transmitted reference views and the synthesized virtual views. To reach the global optimal PAR, a full search was conducted iteratively, with the search step set to 0.05 [29]. Although PAR is not theoretically a discrete variable, simulations proved that the 0.05 step was small enough to indicate any changes. The estimated PAR and the global optimal PAR results for the simulation video sequences are listed in Table 2.

Table 2. Power allocation ratio of the 3D video sequences [29].

Sequence	Estimated PAR	Global Optimal PAR
Balloons	2.23	2.05
Kendo	2.55	2.60
Newspaper	3.32	2.85
Dancer	3.08	3.65
gtFly	5.78	7.00
PoznanHall2	2.84	3.45
PoznanStreet	2.24	2.25
Shark	4.50	4.25

5.2. Minimum Distortion Optimization Performance

We investigated the highest PSNR attained given the available resource (i.e., power and bandwidth). As discussed earlier, we let the maximum number of available channel slots N for one GOP be 8192, where each channel slot transmits only one chunk. As assumed, each GOP transmits a separate texture and depth map for each reference view. Thus the available channel slots N for each GOP are split equally $N_{sub} = \frac{N}{2 \times \text{No. reference views}}$, and N_{sub} becomes the number of available channel slots for each texture and depth map. Furthermore, the noise variance was fixed at 1, and the transmission power budget P was varied for each GOP (Since the noise level is assumed to be fixed, then in the following we no longer differentiate between P and SNR_{chk}).

Figure 2 shows the minimum distortion optimization results for the *Kendo* and *PoznanStreet* sequences. Note that the average results for the two transmitted reference views were used. As shown, for a given bandwidth resource usage N_{sub} , the quality commonly increased with the power budget SNR_{chk} , as was also proved in (12). Nevertheless, for a given power budget SNR_{chk} , the quality did not necessarily increase with the bandwidth usage, mainly under low power budgets. For instance, in the texture frames, when the transmission power P was 5 dB, the PSNR_{max} points for *Kendo* and *PoznanStreet* were achieved when $N_{sub} = 233$ and $N_{sub} = 303$, respectively. Achieving the highest PSNR did not always correspond to the maximum bandwidth usage because, in a pseudo-analog system, different chunks are not equally important, even though each chunk consumes one channel. Hence, more power allocation to HP chunks and less to LP chunks can improve PSNR performance under a limited power budget.

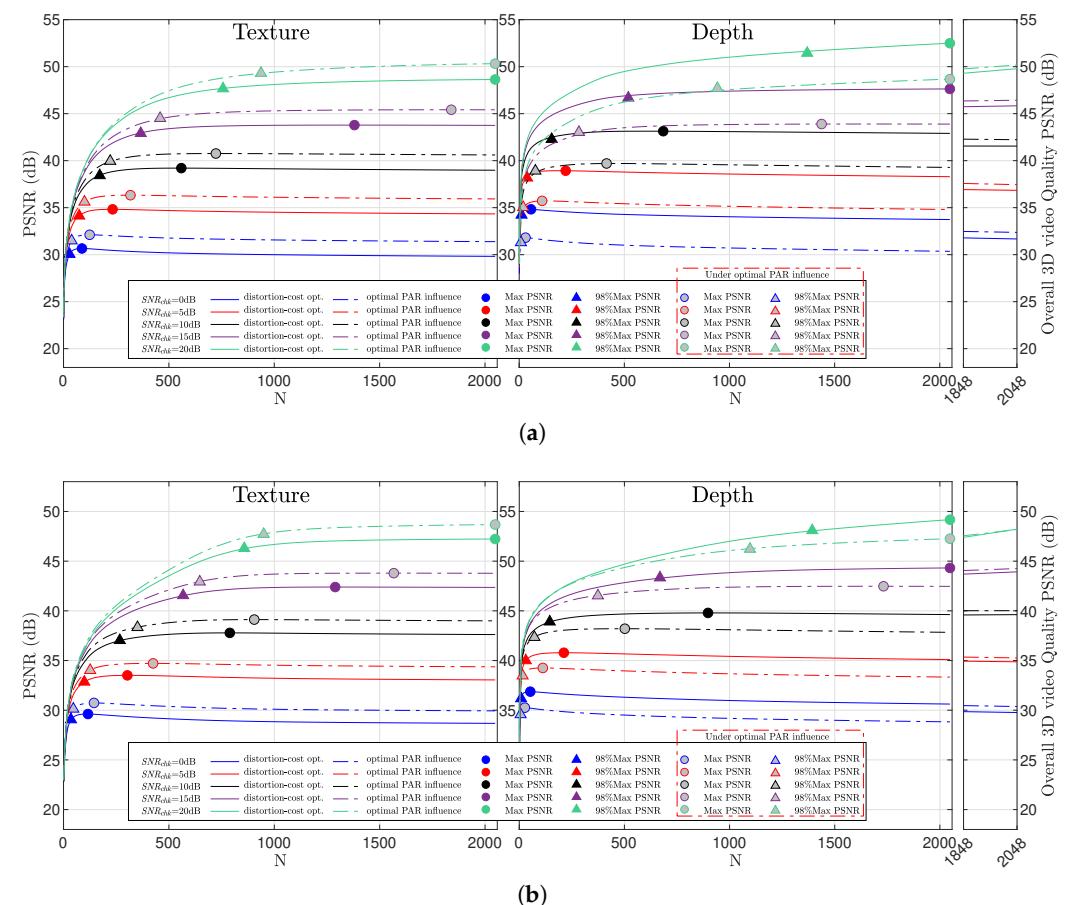


Figure 2. Minimum distortion optimization performance and PAR influence: (a) *Kendo*; (b) *PoznanStreet*.

The minimum distortion optimization curves generally have a flat tail, which implies that above a certain PSNR level, increasing the power budget is more efficient than improving bandwidth use. For instance, for the texture frames in the *Kendo* sequence, under power budget $SNR_{chk} = 10$ dB and bandwidth use $N_{sub} = 559$, the PSNR was 39 dB. The PSNR value did not improve if more bandwidth was assigned. On the other hand, a power increment from 10 dB to 15 dB resulted in a PSNR improvement of 4.3 dB, and a further increment from 10 dB to 20 dB resulted in a PSNR improvement of 7.8 dB. The previous calculations were all carried out under the consideration of an equal texture/depth power allocation ratio or PAR 1:1. The power allocation between the texture and depth map will be rearranged when performing minimum distortion optimization under the optimal PAR mentioned in Table 2. Figure 2 demonstrates the change in behavior after applying the optimal PAR. As more power is assigned to the reference viewpoints' texture, the received quality of the texture frames will be enhanced, and the bandwidth allocation will increase at the expense of the depth map. In many cases, the bandwidth decrease in the depth map is greater than the bandwidth increase in the texture. The overall 3D video quality between two points (1848, 2048) along the bandwidth is drawn to show the improvement in the proposed resource control scheme.

The *Kendo* sequence corresponding to the average performance is plotted in Figure 3 to illustrate the bandwidth usage saving. Both the reference viewpoints' quality, and the five viewpoints' (two reference plus three virtual) overall quality are considered. As a scheme similar to the proposed scheme, the pioneering uncoded video transmission system SoftCast was applied to transmit each reference viewpoint texture and depth map separately, but with equal power allocation between texture and depth map or PAR 1:1. As shown, the conventional SoftCast employs the entire bandwidth and power resources to obtain a good PSNR. However, with less bandwidth usage, the minimum distortion optimization algorithm can obtain a somewhat comparable PSNR before applying the PAR. When the estimated PAR is applied, it alters the total transmission budget SNR_{chk} in (17) according to the optimal ratio, affecting ρ_i^* and K^* for that reference viewpoint texture and depth map and eventually leading to a PSNR enhancement. For instance, in Figure 3d when $SNR_{chk} = 0$ dB, for the overall quality of the 3D video in SoftCast, the bandwidth usage for each reference viewpoint texture and depth map was 2048, and the achieved PSNR was 32.379 dB. However, the minimum distortion algorithm achieved a PSNR of 33.428 dB with only $88/2048 = 4.3\%$ and $56/2048 = 2.7\%$ of the texture and depth bandwidth usage. After applying the optimal PAR, the performance improved to 34.358 dB with a bandwidth usage of 6% and 1.4%. Furthermore, due to the flat tail of the minimum distortion curve, sacrificing a slight margin of the PSNR (e.g., $\tau = 98\%$ of the highest PSNR), reduced the bandwidth usage further from 124 to 39 chunks for texture and from 30 to 8 chunks for the depth map. These figures show that the proposed resource control scheme saves more bandwidth usage at low SNR_{chk} . Hence, it is more suitable when the transmission power budget is limited or under inferior channel conditions.

Table 3 lists the bandwidth usage for different 3D video standards. The distortion-resource and resource-distortion algorithms [31] were modified to ensure a fair comparison. Therefore, instead of transmitting three reference viewpoints, only two were transmitted, and the number of frames in each GOP was increased from four to eight. Since our proposed resource control scheme and the algorithms in [31] exploit different correlation methods, i.e., 3D-DCT and 5D-DCT, the adopted chunk size differs for each scheme, leading to a different number of chunks. Hence, the percentage of the total bandwidth used in each case is calculated, to compare the bandwidth resource usage. The proposed method remarkably decreases the bandwidth usage while achieving satisfactory results.

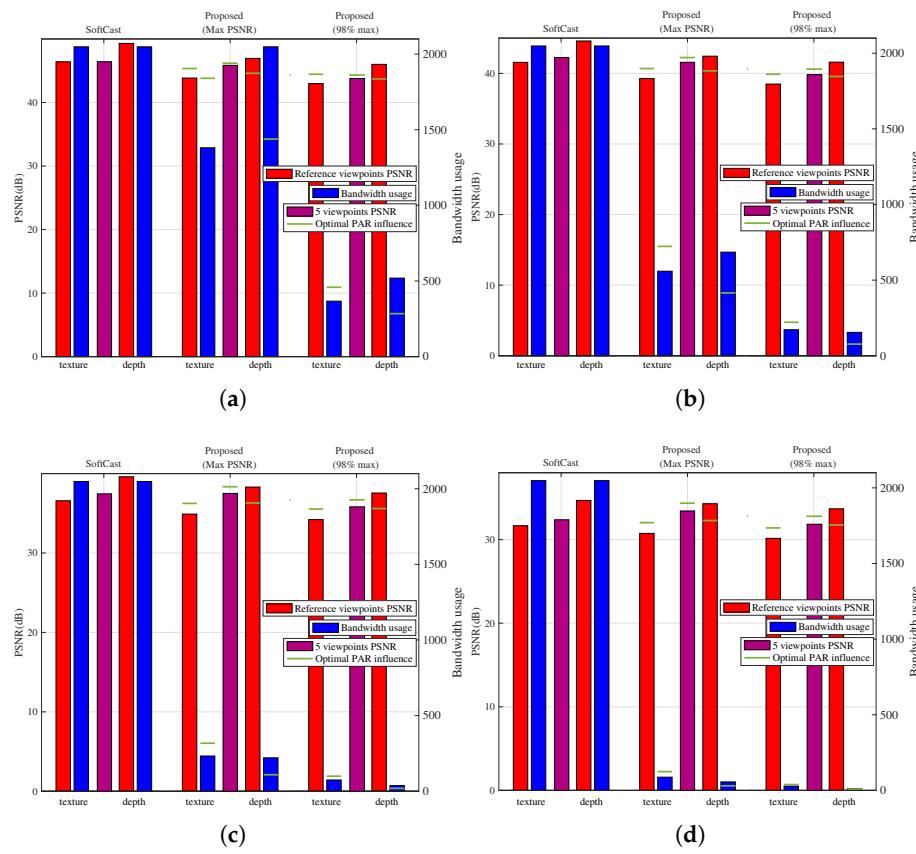


Figure 3. Bandwidth usage savings and PAR influence: (a) $SNR_{chk} = 15 \text{ dB}$; (b) $SNR_{chk} = 10 \text{ dB}$; (c) $SNR_{chk} = 5 \text{ dB}$; (d) $SNR_{chk} = 0 \text{ dB}$.

Table 3. Bandwidth usage comparison for different 3D video sequences.

(a) $SNR_{chk} = 20 \text{ dB}$			
3D Video Sequences	SoftCast	Distortion Resource Algorithm	Proposed Resource Control Scheme
Balloons	100%	67.5292%	51.1108%
Kendo	100%	60.0586%	45.8862%
Newspaper	100%	65.1855%	48.3398%
Dancer	100%	89.3554%	56.4208%
gtFly	100%	91.5039%	49.1577%
PoznanHall2	100%	56.9824%	34.9609%
PoznanStreet	100%	57.0801%	49.9633%
Shark	100%	95.0195%	83.1665%

(b) $SNR_{chk} = 15 \text{ dB}$			
3D Video Sequences	SoftCast	Distortion Resource Algorithm	Proposed Resource Control Scheme
Balloons	100%	36.9629%	19.5922%
Kendo	100%	30.6641%	18.1274%
Newspaper	100%	33.8379%	17.6775%
Dancer	100%	73.3398%	40.8203%
gtFly	100%	75.0977%	41.6748%
PoznanHall2	100%	37.5000%	21.6919%
PoznanStreet	100%	34.6191%	24.9146%
Shark	100%	84.9609%	59.1309%

Table 3. Cont.

(c) $SNR_{chk} = 10$ dB			
3D Video Sequences	SoftCast	Distortion Resource Algorithm	Proposed Resource Control Scheme
Balloons	100%	9.3262%	4.5898%
Kendo	100%	12.0117%	7.2875%
Newspaper	100%	10.4980%	4.3212%
Dancer	100%	40.3809%	22.4365%
gtFly	100%	39.2578%	25.8056%
PoznanHall2	100%	19.5801%	11.7187%
PoznanStreet	100%	17.2852%	10.3271%
Shark	100%	49.7558%	19.2138%
(d) $SNR_{chk} = 5$ dB			
3D Video Sequences	SoftCast	Distortion Resource Algorithm	Proposed Resource Control Scheme
Balloons	100%	2.7344%	1.5380%
Kendo	100%	4.7852%	2.8930%
Newspaper	100%	3.4180%	1.5014%
Dancer	100%	11.9141%	6.8603%
gtFly	100%	8.9356%	8.0688%
PoznanHall2	100%	5.8594%	3.4058%
PoznanStreet	100%	6.3965%	3.4790%
Shark	100%	12.3046%	7.0313%
(e) $SNR_{chk} = 0$ dB			
3D Video Sequences	SoftCast	Distortion Resource Algorithm	Proposed Resource Control Scheme
Balloons	100%	1.0254%	0.6225%
Kendo	100%	2.1484%	1.1352%
Newspaper	100%	1.1230%	0.6225%
Dancer	100%	2.5391%	1.5136%
gtFly	100%	1.2207%	1.3672%
PoznanHall2	100%	2.2461%	0.9521%
PoznanStreet	100%	2.2539%	1.2939%
Shark	100%	3.4179%	2.0019%

5.3. Target Distortion Optimization Performance

5.3.1. Power Distortion Optimization Performance

Assume the sender delivers about 30 GOPs to users whose demanded quality is 35 dB, given the bandwidth constraint $n = 8192$ chunks per GOP. Figure 4 depicts the power allocation for the consecutive GOPs and the total bandwidth usage for *Kendo* and *PoznanStreet* sequence reference viewpoints. It is apparent that the power allocation for depth map video is lower than for texture due to depth map's because of depth map distinct characteristics (i.e., the 3D scene's geometric information which comes in the form of homogeneous regions separated by sharp edges). The same is true for bandwidth usage. In some GOPs, depth maps may occupy less than 0.005% of the available bandwidth, as in the *PoznanStreet* sequence. It is evident that the power allocation curves for both texture and depth map frames are relatively smooth, and the total bandwidth usage is kept at a comparatively low level. This complex resource allocation assists in maintaining the 3D video quality at the predefined PSNR value. Thus, the content is delivered with a favorable viewing experience at the user end, and other users can also benefit from the saved resources.

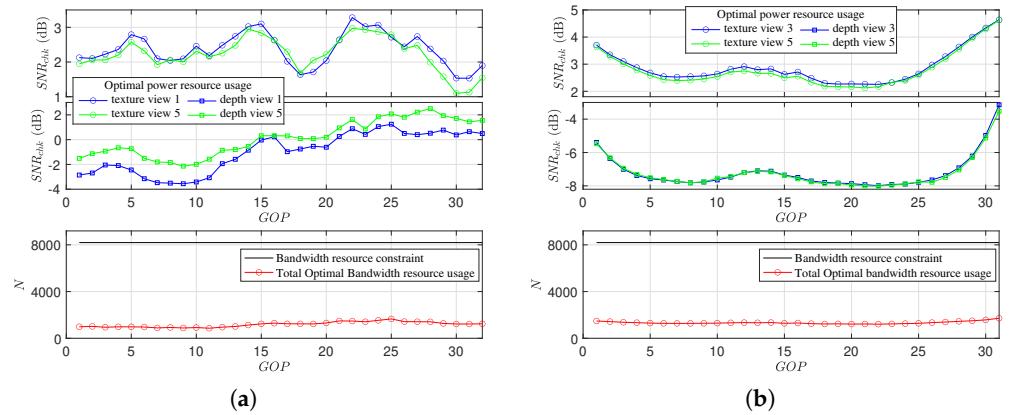


Figure 4. Power allocation for texture and depth map of 3D video sequences: (a) Kendo; (b) PoznanStreet.

5.3.2. Bandwidth Distortion Optimization Performance

Similarly, assume the sender delivers about 30 GOPs to users whose demanded quality is 35 dB, under a transmission power constraint $SNR_{clk} = 10$ dB. Figure 5 shows the bandwidth allocation curve for reference viewpoints for different video sequences. Compared with the large chunk number $n = 8192$ in one GOP, the total bandwidth usage fluctuates at a low and steady level. The bandwidth usage for the depth map can be less than three chunks in some GOPs. Therefore, it will be practical to allocate slightly more bandwidth resources (i.e., 4%), to avoid running Algorithm 3 several times, which will reduce the computational cost.

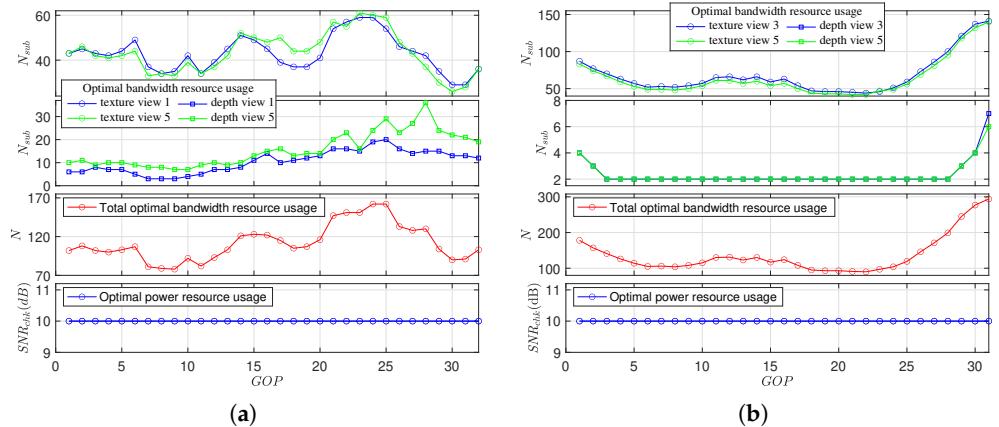


Figure 5. Bandwidth usage for texture and depth map of 3D video sequences: (a) Kendo; (b) PoznanStreet.

Figure 6 presents the bandwidth usage saving obtained while achieving the target PSNRs. The conventional SoftCast exhausts all the available resources to achieve a good PSNR, unaware of the user's predefined distortion. However, the target distortion algorithm approaches the user's predefined PSNR with a lower bandwidth usage. The performance is slightly enhanced when the optimal PAR is applied. Based on Algorithms 2 and 3, the optimal PAR does not influence the power and bandwidth allocation. Nonetheless, provided the optimal PAR adjusts the power assigned jointly to texture and depth map frames in the pseudo-analog transmission, as shown in Figure 1, it still can influence the overall quality of the 3D video at the receiving end.

For example, in Figure 6c, when the power budget $SNR_{clk} = 10$ dB and the target PSNR = 40 dB, in SoftCast, the bandwidth usage for texture and depth map videos with a single reference viewpoint is N_{sub} and N_{sub} , respectively, and the achieved overall quality

of 3D video is 42.26 dB. However, we achieved a PSNR of 39.30 dB with the proposed target distortion algorithm using only $167/N_{sub} = 8.15\%$ and $42/N_{sub} = 2\%$ of the original bandwidth use for the texture and depth map. When the optimal PAR was applied, the proposed resource control scheme achieved a PSNR of 39.66 dB, close to the desired quality, while maintaining the same bandwidth usage.

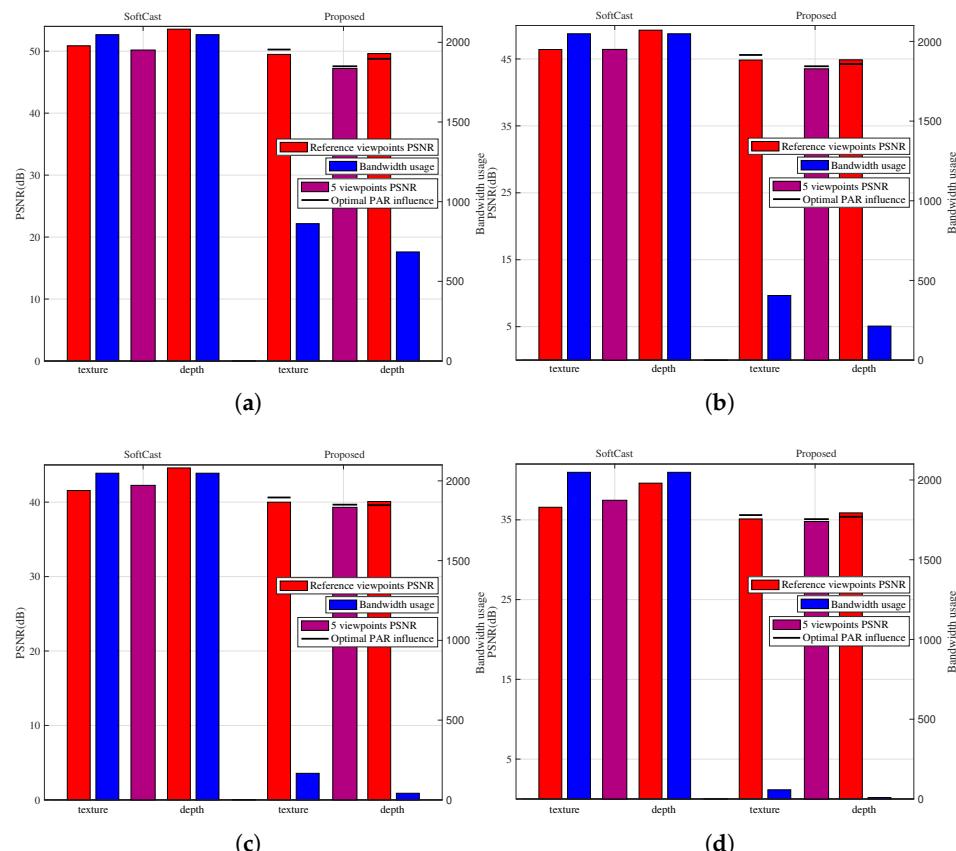


Figure 6. Bandwidth usage savings and PAR influence: (a) $SNR_{chk} = 20$ dB ($PSNR_{target} = 50$ dB); (b) $SNR_{chk} = 15$ dB ($PSNR_{target} = 45$ dB); (c) $SNR_{chk} = 10$ dB ($PSNR_{target} = 40$ dB); (d) $SNR_{chk} = 5$ dB ($PSNR_{target} = 35$ dB).

The bandwidth usage saving while achieving a target PSNR under a power budget $SNR_{chk} = 20$ dB is shown in Figure 7. As mentioned earlier, the bandwidth resource usage percentage will be used for the comparison between our proposed scheme and the resource-distortion algorithm [31]. In this figure, the total bandwidth usage percentage is split equally between the texture and depth map. The proposed resource control scheme achieves PSNR results closer to the target quality than the resource-distortion algorithm, with less bandwidth usage. For example, in Figure 7b, the target PSNR = 45 dB. For the resource-distortion algorithm, the bandwidth usage to achieve 43.379 dB was $149/512 = 7.25\%$ for texture and the same for the depth map. In contrast, the proposed resource control scheme achieved a PSNR of 43.556 dB with only $558/2048 = 6.8\%$ and $350/2048 = 4.25\%$ bandwidth usage for the texture and depth map, respectively.

5.3.3. Power and Bandwidth Trade-Off

A trade-off exists between power and bandwidth usage for each expected video quality, as the minimal resource solution is generally not unique. However, the valid ranges for power and bandwidth for a given distortion are constrained. The trade-off curves shown in Figure 8 correspond to the average texture and depth maps. The curve boundaries vary with different predefined distortions. In general, for lower target PSNR values, the required power and bandwidth usages are comparatively minimal. The power resource is high

near the N_{sub} lower boundary. At this point, a slight increase in bandwidth dramatically reduces the power. Nonetheless, allocating more bandwidth after a certain threshold does not reduce the power further. Hence, it is suggested that a minimum power and bandwidth usage pair is adopted close to the curve's turning point, where a slight change in one factor noticeably affects the other.

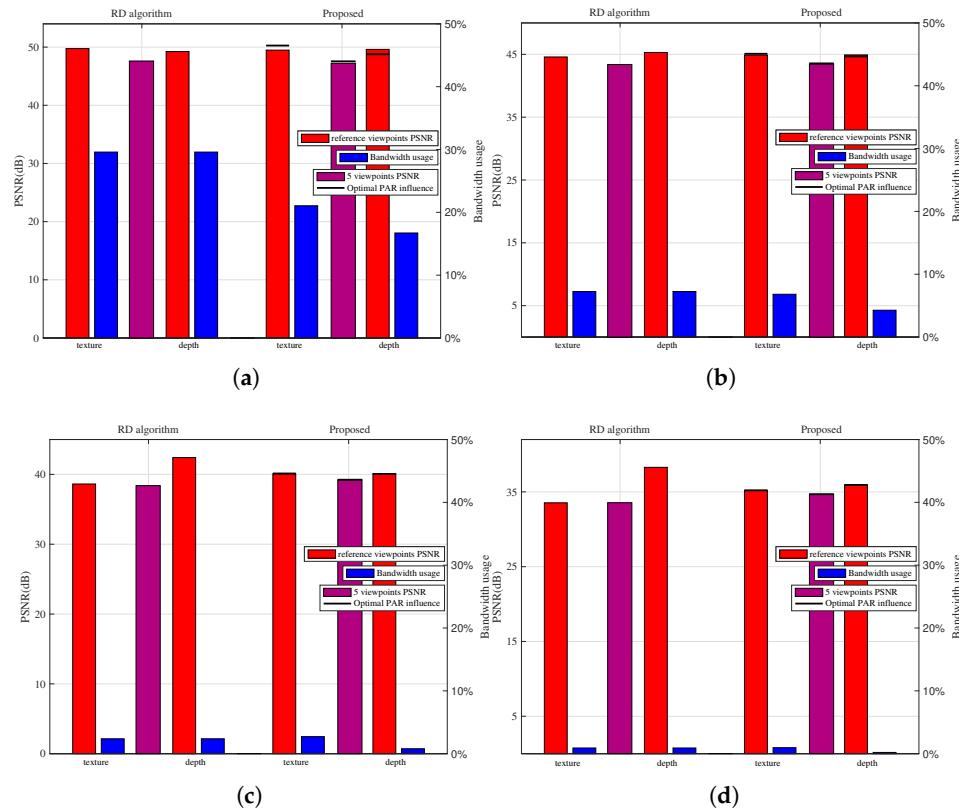


Figure 7. Bandwidth usage savings and PAR influence under $SNR_{chk} = 20$ dB: (a) $PSNR_{target} = 50$ dB; (b) $PSNR_{target} = 45$ dB; (c) $PSNR_{target} = 40$ dB; (d) $PSNR_{target} = 35$ dB.

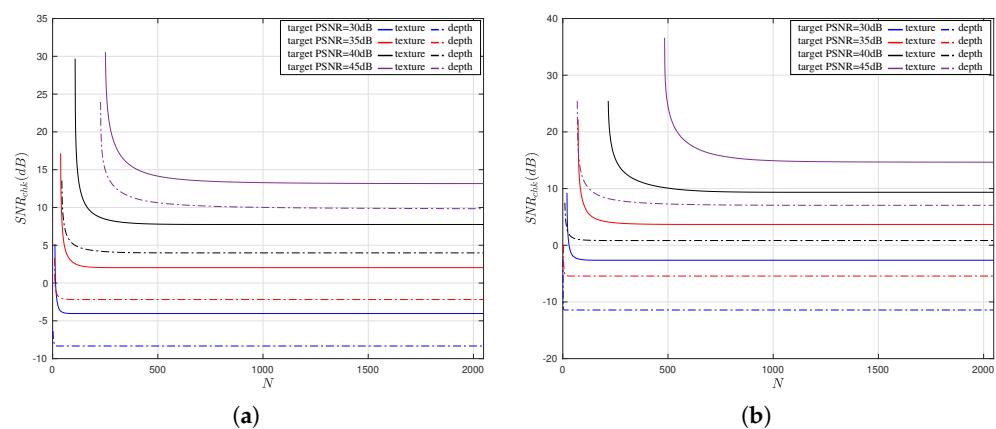


Figure 8. Power and bandwidth usage trade-off curves for different video sequences: (a) Kendo; (b) PoznanStreet.

5.4. The Resource Control Scheme Performance Comparison

The performance of the proposed scheme was tested regarding the synthesized virtual views quality and compared with two benchmark schemes: SoftCast and the distortion-resource algorithm [31].

Figure 9 illustrates the synthesis quality at virtual viewpoint 2 of the 3D video for the three schemes, along with the synthesized frames. Under a channel bandwidth constraint of 8192, the SoftCast scheme exploits all the available bandwidth resources and allocates power to each chunk. In contrast, the other schemes tend to retain HP chunks and discard LP chunks, which significantly saves bandwidth usage. For instance, when the power usage was 10 dB, the PSNR and SSIM of the proposed resource control scheme were close to the values obtained by SoftCast with only 7.28% bandwidth usage, which is nearly half of the bandwidth saved by the distortion-resource algorithm. As the power usage decreased to SNR = 0 dB, the proposed scheme achieved even better performance. With only 1.1535% bandwidth usage, the proposed scheme PSNR saved 1.21 dB over SoftCast, and in terms of SSIM, the proposed scheme improved the SSIM from 0.8244 to 0.9317.

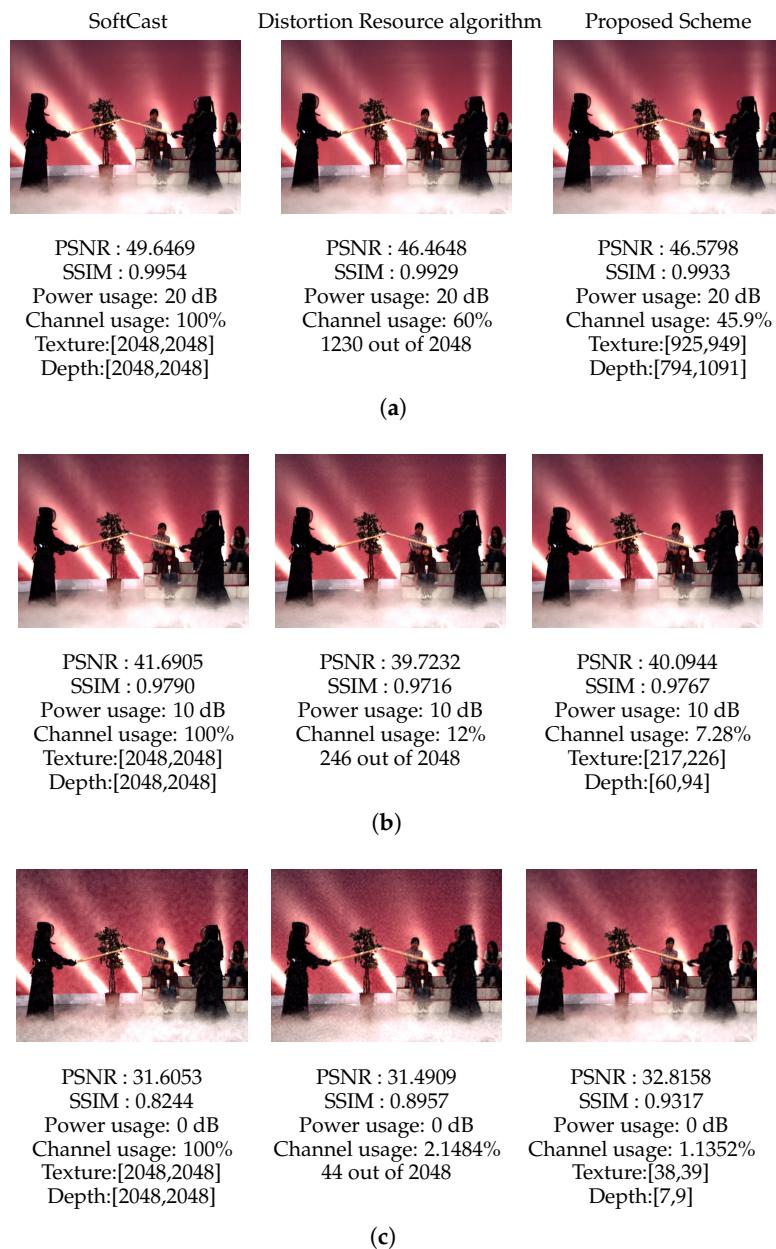


Figure 9. Synthesis quality at virtual view point 2 for different algorithms: (a) SNR = 20 dB; (b) SNR = 10 dB; (c) SNR = 0 dB.

The robustness of the proposed scheme was evaluated by a comparison of the results of the received video quality with those of the conventional SoftCast and the distortion-

resource algorithm. The performance achieved by each method in Table 4 corresponds to the bandwidth usage illustrated in Table 3. SoftCast harnesses all the bandwidth and assigns the available power to all chunks. The distortion-resource algorithm utilizes the available power by assigning the power resource to an optimal number of chunks, to minimize the receiving viewpoints' distortion. In contrast, the proposed scheme optimally allocates the available power to a limited number of chunks such that the overall quality of the transmitted 3D video is maximized. For instance, for the low power usage of 0 dB, the proposed method could obtain a PSNR gain of 1.5 dB over SoftCast and even a 1.789 dB gain over the distortion-resource algorithm (gtFly sequence), using less than 1.4% of the bandwidth, and a maximum SSIM increment of 0.1588 was achieved over SoftCast (Shark sequence) using 2% of the bandwidth.

Table 4. Preformance comparison for different 3D video sequences.

(a) $SNR_{chk} = 20 \text{ dB}$						
3D Video Sequences	SoftCast		Distortion Resource Algorithm		Proposed Resource Control Scheme	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Balloons	49.3196	0.9964	46.1196	0.9942	46.2761	0.9943
Kendo	49.7899	0.9957	47.6392	0.9936	48.0136	0.9939
Newspaper	49.2141	0.9963	46.6435	0.9943	46.9753	0.9946
Dancer	46.0253	0.9916	43.2292	0.9893	43.6010	0.9913
gtFly	48.9609	0.9930	47.6194	0.9916	47.1173	0.9927
PoznanHall2	53.1829	0.9969	50.1385	0.9944	51.2116	0.9957
PoznanStreet	48.2154	0.9952	45.1048	0.9912	45.9560	0.9925
Shark	48.1204	0.9933	46.3205	0.9899	47.0311	0.9929
(b) $SNR_{chk} = 15 \text{ dB}$						
3D Video Sequences	SoftCast		Distortion Resource Algorithm		Proposed Resource Control Scheme	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Balloons	45.4115	0.9928	42.5971	0.9878	42.5716	0.9878
Kendo	45.8277	0.9911	43.9952	0.9862	44.3260	0.9875
Newspaper	45.1731	0.9921	42.8919	0.9877	43.0334	0.9884
Dancer	41.5428	0.9781	39.2760	0.9740	39.6065	0.9790
gtFly	45.1369	0.9845	43.6927	0.9814	43.8269	0.9852
PoznanHall2	49.0760	0.9935	46.2265	0.9882	47.1799	0.9905
PoznanStreet	43.9320	0.9888	41.5915	0.9812	42.2585	0.9837
Shark	44.1887	0.9845	42.5072	0.9776	42.9480	0.9847
(c) $SNR_{chk} = 10 \text{ dB}$						
3D Video Sequences	SoftCast		Distortion Resource Algorithm		Proposed Resource Control Scheme	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Balloons	41.3268	0.9830	39.2838	0.9743	39.3090	0.9769
Kendo	41.5470	0.9780	39.8723	0.9696	40.6042	0.9760
Newspaper	40.7441	0.9801	39.0226	0.9728	38.7993	0.9733
Dancer	37.0933	0.9402	35.0093	0.9329	35.3295	0.9453
gtFly	40.6633	0.9588	39.0618	0.9545	40.3215	0.9655
PoznanHall2	44.5969	0.9844	42.3917	0.9731	43.6213	0.9788
PoznanStreet	39.5928	0.9713	37.7700	0.9557	38.5942	0.9634
Shark	39.7686	0.9587	38.5368	0.9516	38.6494	0.9663

Table 4. Cont.

		(d) $SNR_{chk} = 5 \text{ dB}$					
3D Video Sequences		SoftCast		Distortion Resource Algorithm		Proposed Resource Control Scheme	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Balloons	36.8079	0.9526	35.8347	0.9512	36.2920	0.9613	
Kendo	36.8431	0.9389	35.5959	0.9398	36.6900	0.9568	
Newspaper	36.1632	0.9460	35.0736	0.9391	35.3860	0.9479	
Dancer	32.4094	0.8434	30.9076	0.8469	31.6366	0.8695	
gtFly	36.0398	0.8898	34.4277	0.9028	36.0597	0.9268	
PoznanHall2	40.0081	0.9584	38.4855	0.9466	39.8148	0.9566	
PoznanStreet	34.8735	0.9192	33.8732	0.9069	34.8194	0.9270	
Shark	35.0505	0.8880	34.0345	0.9067	34.9220	0.9307	

		(e) $SNR_{chk} = 0 \text{ dB}$					
3D Video Sequences		SoftCast		Distortion Resource Algorithm		Proposed Resource Control Scheme	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Balloons	31.9165	0.8578	32.0593	0.9015	32.9264	0.9328	
Kendo	31.6785	0.8227	31.3943	0.8887	32.7896	0.9280	
Newspaper	31.1809	0.8513	30.9788	0.8707	32.0076	0.9064	
Dancer	27.4676	0.6483	27.8362	0.7346	28.5260	0.7732	
gtFly	31.1336	0.7310	30.8460	0.8469	32.6345	0.8842	
PoznanHall2	35.0796	0.8827	34.7832	0.9127	36.2844	0.9345	
PoznanStreet	29.7831	0.7757	30.2480	0.8329	31.2267	0.8715	
Shark	29.9380	0.7188	30.2311	0.8336	31.2955	0.8776	

Figure 10 shows the constrained resource influence on pseudo-analog transmission. Assuming different bandwidth resources along with different power budgets reveals the pseudo-analog transmission behavior. The impact of discarding LP chunks and retaining HP chunks at lower SNR is different from that at higher SNR. For each given power budget, several bandwidth constraints are suggested. Under each bandwidth resource, SoftCast consumes the available bandwidth resource for the transmission process. Although the available bandwidth resource also bounds the proposed resource control scheme, it achieves $PSNR_{max}$. To save more bandwidth, the proposed scheme was set to achieve $\tau \cdot PSNR_{max}$. The dashed line represents the resulting video quality for SoftCast transmission utilizing the whole bandwidth. In Figure 10a the same result is shown as in Figure 9 but for only one virtual viewpoint.

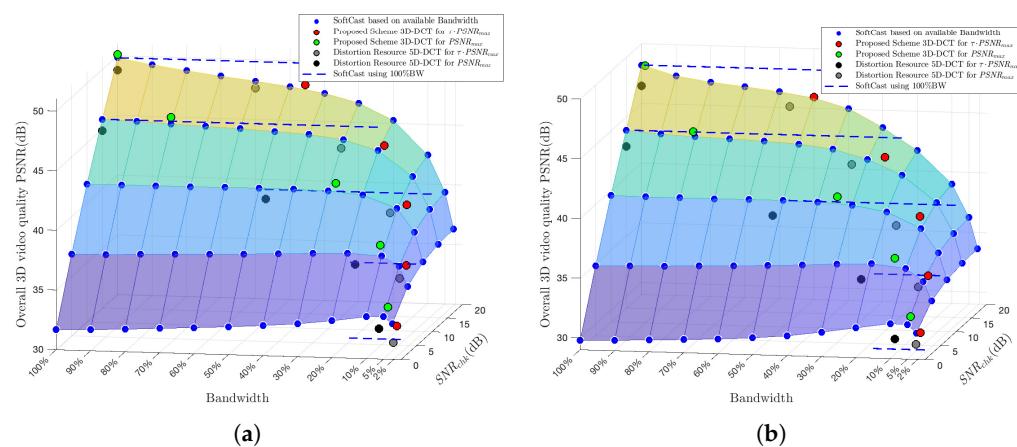


Figure 10. Performance comparison of different schemes under multiple SNR_{chk} and bandwidth values: (a) Kendo; (b) PoznanStreet.

5.5. Complexity Analysis

The proposed resource control scheme complexity is divided into two parts.

In the joint power allocation method, the complexity is almost determined by the preprocessing step. The full search method complexity is directly proportional to the PAR search range. If a search range of R is assumed with a search step of 0.05, the search will preprocess $20 \times R$ times, including $20 \times R \times M$ times distortion computing of reference viewpoints and $20 \times R \times L$ times virtual view synthesis. However, the estimation method will preprocess only 6 times, including 6×2 times distortion computing of the leftmost and rightmost reference views and $6 \times L$ times virtual view synthesis. Only the first frame of each sequence is preprocessed in the proposed estimation method, whereas the whole sequence is preprocessed in the full search method. The significance becomes clear as the sequence frames number grows. Thus, the estimation method complexity is 96.25% lower than that of the full search method, which is trivial.

In the greedy search algorithms (Algorithms 1–3), all possible chunk (n) or channel (r) values are traversed iteratively, and for each n or r , all the operations involved are linear. Thus, Algorithms 1–3 all have an $O(N)$ complexity, which is negligible. The N chunks energy sorting process in each GOP is the major source of complexity, and is $O(N \log N)$. However, instead of strictly sorting all the chunks according to their energy distribution, the sorting process can be avoided, as chunks can be sorted in a zigzag scanning manner (i.e., JPEG image compression). Furthermore, the complexity can be further reduced by considering the calculation of all the optimal values for the first GOP only and then applying the result to the remaining GOPs. This is possible if the video content of consecutive frames does not change greatly.

6. Conclusions

This paper presented a resource control scheme for pseudo-analog transmission based on the integration of joint power allocation with resource optimization for multi-view plus depth video transmission. While the joint texture/depth power allocation maximized the transmission efficiency, the resource allocation algorithms handled the data traffic burden caused by multi-view video content. The estimated optimal PAR between texture and depth performed similarly to the full search method but with negligible complexity. Following the SoftCast perspective, a minimum distortion algorithm was used to achieve the best viewing quality under a resource constraint. In contrast, the target distortion algorithm minimizes the resource usage for a predefined distortion requirement, which results in receiving videos of consistent viewing quality. As the minimal resource solution is generally not unique, we analyzed the power and bandwidth usage trade-off. The reported results verified the efficiency of the proposed scheme for reference and virtual viewpoints. During 3D video streaming, if the 3D scene changes, the video content characteristic in the subsequent frames may differ. Hence, the model parameters of the joint power allocation estimation using the first frames become inaccurate for the other frames, which have video content with considerably different characteristics. Adding scene change detection allows the system to update the model parameters automatically whenever the video content changes dramatically. In future work, the inter-view correlation could be exploited to optimize the transmission performance using 4D-DCT.

Author Contributions: Conceptualization, S.K.S.T., O.A., A.M.S., E.O.-M. and O.B.; methodology, S.K.S.T., O.B. and O.A.; software, S.K.S.T., O.A. and E.O.-M.; validation, S.K.S.T., O.A. and E.O.-M.; formal analysis, S.K.S.T. and O.B.; investigation, S.K.S.T.; resources, S.K.S.T. and O.A.; data curation, S.K.S.T. and O.A.; writing—original draft preparation, S.K.S.T. and E.O.-M.; writing—review and editing, S.K.S.T., E.O.-M. and O.B.; visualization, S.K.S.T., A.M.S. and E.O.-M.; supervision, S.K.S.T. and O.B.; project administration, S.K.S.T., O.B. and E.O.-M.; funding acquisition, S.K.S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge the support of our laboratory, the Advanced Visual Communication and Computing (AVC2) laboratory.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Merkle, P.; Smolic, A.; Muller, K.; Wiegand, T. Multi-view video plus depth representation and coding. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16 September–19 October 2007; Volume 1, pp. 201–204.
2. Cao, B.; Xia, S.; Han, J.; Li, Y. A distributed game methodology for crowdsensing in uncertain wireless scenario. *IEEE Trans. Mob. Comput.* **2020**, *19*, 15–28. [[CrossRef](#)]
3. Fehn, C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In Proceedings of the Stereoscopic Displays and Virtual Reality Systems XI, San Jose, CA, USA, 19–22 January 2004; Volume 5291, pp. 93–104.
4. Zhu, C.; Li, S. Depth image based view synthesis: New insights and perspectives on hole generation and filling. *IEEE Trans. Broadcast.* **2016**, *62*, 82–93. [[CrossRef](#)]
5. Li, S.; Zhu, C.; Sun, M.T. Hole filling with multiple reference views in DIBR view synthesis. *IEEE Trans. Multimed.* **2018**, *20*, 1948–1959. [[CrossRef](#)]
6. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
7. Vembu, S.; Verdu, S.; Steinberg, Y. The source-channel separation theorem revisited. *IEEE Trans. Inf. Theory* **1995**, *41*, 44–54. [[CrossRef](#)]
8. Yuan, H.; Fu, H.; Liu, J.; Xiao, J. End-to-end distortion-based multiuser bandwidth allocation for real-time video transmission over LTE network. *IEEE Trans. Broadcast.* **2017**, *63*, 338–349. [[CrossRef](#)]
9. Kratochvíl, T.; Štukavec, R. DVB-T digital terrestrial television transmission over fading channels. *Radioengineering* **2008**, *17*, 96–102.
10. Katahi, D.; Rahul, H.; Jakubczak, S. *Softcast: One Video to Serve All Wireless Receivers*; Tech. Rep. MIT-CSAIL-TR-2009-005; Massachusetts Institute of Technology: Cambridge, MA, USA, 2009.
11. Jakubczak, S.; Katahi, D. SoftCast: One-size-fits-all wireless video. In Proceedings of the ACM SIGCOMM 2010 Conference, New Delhi, India, 30 August–3 September 2010; pp. 449–450.
12. Jakubczak, S.; Katahi, D. A cross-layer design for scalable mobile video. In Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, Las Vegas, NV, USA, 19–23 September 2011; pp. 289–300.
13. Giannopoulos, A.; Spantideas, S.; Kapsalis, N.; Karkazis, P.; Trakadas, P. Deep reinforcement learning for energy-efficient multi-channel transmissions in 5G cognitive hetnets: Centralized, decentralized and transfer learning based solutions. *IEEE Access* **2021**, *9*, 129358–129374. [[CrossRef](#)]
14. Cumino, P.; Lobato Junior, W.; Tavares, T.; Santos, H.; Rosário, D.; Cerqueira, E.; Villas, L.A.; Gerla, M. Cooperative UAV scheme for enhancing video transmission and global network energy efficiency. *Sensors* **2018**, *18*, 4155. [[CrossRef](#)]
15. Fan, X.; Xiong, R.; Wu, F.; Zhao, D. WaveCast: Wavelet based wireless video broadcast using lossy transmission. In Proceedings of the 2012 Visual Communications and Image Processing, San Diego, CA, USA, 27–30 November 2012; pp. 1–6.
16. Fan, X.; Wu, F.; Zhao, D.; Au, O.C. Distributed wireless visual communication with power distortion optimization. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1040–1053. [[CrossRef](#)]
17. Xiong, R.; Wu, F.; Fan, X.; Luo, C.; Ma, S.; Gao, W. Power-distortion optimization for wireless image/video SoftCast by transform coefficients energy modeling with adaptive chunk division. In Proceedings of the 2013 Visual Communications and Image Processing (VCIP), Kuching, Malaysia, 17–20 November 2013; pp. 1–6.
18. Cui, H.; Xiong, R.; Luo, C.; Song, Z.; Wu, F. Denoising and resource allocation in uncoded video transmission. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 102–112. [[CrossRef](#)]
19. Balsa, J.; Fresnedo, Ó.; García-Naya, J.A.; Domínguez-Bolaño, T.; Castedo, L. JSCC-Cast: A Joint Source Channel Coding Video Encoding and Transmission System with Limited Digital Metadata. *Sensors* **2021**, *21*, 6208. [[CrossRef](#)]
20. Cui, H.; Song, Z.; Yang, Z.; Luo, C.; Xiong, R.; Wu, F. Cactus: A hybrid digital-analog wireless video communication system. In Proceedings of the 16th ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems, Barcelona, Spain, 3–8 November 2013; pp. 273–278.
21. Liu, X.L.; Hu, W.; Luo, C.; Pu, Q.; Wu, F.; Zhang, Y. ParCast+: Parallel video unicast in MIMO-OFDM WLANs. *IEEE Trans. Multimed.* **2014**, *16*, 2038–2051. [[CrossRef](#)]
22. He, D.; Luo, C.; Lan, C.; Wu, F.; Zeng, W. Structure-preserving hybrid digital-analog video delivery in wireless networks. *IEEE Trans. Multimed.* **2015**, *17*, 1658–1670. [[CrossRef](#)]
23. Yu, L.; Li, H.; Li, W. Wireless scalable video coding using a hybrid digital-analog scheme. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 331–345. [[CrossRef](#)]

24. Yu, L.; Li, H.; Li, W. Wireless cooperative video coding using a hybrid digital–analog scheme. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 436–450. [[CrossRef](#)]
25. Liang, F.; Luo, C.; Xiong, R.; Zeng, W.; Wu, F. Superimposed Modulation for Soft Video Delivery with Hidden Resources. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2345–2358. [[CrossRef](#)]
26. He, C.; Wang, H.; Hu, Y.; Chen, Y.; Fan, X.; Li, H.; Zeng, B. MCast: High-quality linear video transmission with time and frequency diversities. *IEEE Trans. Image Process.* **2018**, *27*, 3599–3610. [[CrossRef](#)]
27. He, C.; Hu, Y.; Chen, Y.; Fan, X.; Li, H.; Zeng, B. Mucast: Linear uncoded multiuser video streaming with channel assignment and power allocation optimization. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 1136–1146. [[CrossRef](#)]
28. Yang, T.; Luo, L.; Zhu, C.; Tang, S. Block DCT based optimization for wireless SoftCast of depth map. *IEEE Access* **2019**, *7*, 29484–29494. [[CrossRef](#)]
29. Luo, L.; Yang, T.; Zhu, C.; Jin, Z.; Tang, S. Joint texture/depth power allocation for 3-D video SoftCast. *IEEE Trans. Multimed.* **2019**, *21*, 2973–2984. [[CrossRef](#)]
30. Liu, D.; Wu, J.; Cui, H.; Zhang, D.; Luo, C.; Wu, F. Cost-distortion optimization and resource control in pseudo-analog visual communications. *IEEE Trans. Multimed.* **2018**, *20*, 3097–3110. [[CrossRef](#)]
31. Zhang, T.; Mao, S. Joint power and channel resource optimization in soft multi-view video delivery. *IEEE Access* **2019**, *7*, 148084–148097. [[CrossRef](#)]
32. Xiong, R.; Wu, F.; Xu, J.; Fan, X.; Luo, C.; Gao, W. Analysis of decorrelation transform gain for uncoded wireless image and video communication. *IEEE Trans. Image Process.* **2016**, *25*, 1820–1833.
33. Cui, H.; Luo, C.; Chen, C.W.; Wu, F. Robust uncoded video transmission over wireless fast fading channel. In Proceedings of the IEEE INFOCOM 2014–IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 73–81.
34. Saito, T. Nagoya University Multi-View Sequences Download List. Nagoya University, Fujii Laboratory. 2015. Available online: <http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/> (accessed on 1 May 2015).
35. Domąski, M.; Grajek, T.; Klimaszewski, K.; Kurc, M.; Stankiewicz, O.; Stankowski, J.; Wegner, K. Poznan multiview video test sequences and camera parameters. Paper presented at the MPEG/M17050, Xian, China, October 2009. ISO/IEC JTC1/SC29/WG11.
36. Fraunhofer Heinrich Hertz Institute. 3D High Efficiency Video Coding (3D-HEVC)—JCT-VC. Available online: <https://hevc.hhi.fraunhofer.de/3dhevc> (accessed on 10 September 2021).